# Center for Independent Experts (CIE) Independent Peer Review Report on the 2015 Assessment of Eastern Bering Sea Pollock

by R.I.C. Chris Francis
123 Overtoun Terrace
Wellington 6021
New Zealand

July 2016

Prepared for

The Center for Independent Experts

# Executive Summary

A stock assessment review panel met 16-19 May 2016 at the Alaska Fisheries Science Center of NOAA/NMFS in Seattle to review the 2015 assessment of the East Bering Sea stock of walleye pollock (*Gadus chalcogrammus*). The assessment and related material were presented to the Panel and some additional analyses were requested and considered.

The review was very well run and I conclude that the assessment is the best scientific information available for this stock.

As clear improvements for future assessments I recommend that:

1. The index from the ATS survey include echoes down to the top of the acoustic dead zone (Section 3.1.0)
2. The body-mass growth model be modified to produce more reliable predictions of future values (Section 3.1.3)

As refinements to future assessments (and the Tier system) I recommend that:

3. The use of age-length keys in constructing age composition data be made more consistent (Section 3.1.0)
4. Consideration be given to extending the current otolith-reading quality-control procedures to include an explicit check of the readers' decision to read some otoliths whole (Section 3.1.0)
5. The BTS and ATS-age-2+ surveys be treated as indexing biomass if it is confirmed that these surveys estimate biomass more precisely than numbers of fish (Section 3.1.2)
6. The weighting of data inputs in the assessment model be done more rigorously (Section 3.1.4)
7. The current approach to selectivity be evaluated by (a) formally comparing it to alternative approaches to ensure that it is adequately flexible but not excessively complex; (b) investigating whether it might be improved by disaggregating the fishery data (to provide more direct information about year-to-year changes in selectivity); and (c) testing to ensure that the composition data are not over-fitted (Section 3.2.1)
8. The stock-recruit steepness parameter be fixed in the assessment model (Section 3.2.5)
9. The ATS-age-1 index be dropped from the assessment (Section 3.2.5)
10. Consideration be given to making the criterion for a fish stock's inclusion in Tier 1 less unrealistically stringent (Section 3.3)

To facilitate future reviews of this assessment I recommend that:

11. The assessment documentation be improved by (a) ensuring that it is consistent with the current assessment and (b) including as appendices the model code (i.e., the tpl) and input files (Section 3.5)

# 1. Background

This report reviews, at the request of the Center for Independent Experts (CIE) (see Appendix 2), the 2015 assessment for walleye pollock (*Gadus chalcogrammus*) in the East Bering Sea. The fishery for this stock, which is very large (with average catch about 1.2 million tonnes per year), is managed by the North Pacific Management Council within their Bering Sea and Aleutian Islands Management Plan.

# 2. Reviewer's Role

The role of the present reviewer was to read the documents provided (see Appendix 1), participate in the panel/peer review meeting held at the Alaska Fisheries Science Center in Seattle, WA during May 16-19, 2016 (see Appendix 3 for a list of participants), and write an independent peer review report in accordance with his Terms of Reference (given in Annex 2 of Appendix 2).

# 3. Findings

I was impressed by the running of the review meeting. Both the stock assessment and the associated management system are relatively complex but every effort was made – both by the chair and presenters – to respond to queries by the reviewers. The only weakness in this respect was in the documentation of the assessment (see Section 3.5).

My findings are grouped in accordance with the Terms of References (TORs) for the review, as given in Annex 2 of Appendix 2. In what follows, "AR" will be used to identify references to parts (e.g., sections, tables, figures) of the Assessment Report.

## 3.1 TOR 1: Quality and processing of the input data

**Evaluation, findings, and recommendations on quality of input data and methods used to process them for inclusion in the assessment.**

Overall, I found the input data for this assessment to be of a high standard. I was impressed with the thoroughness of the catch sampling and otolith reading programmes and the design, implementation, and analysis of the surveys.

Before addressing the specific topics of TOR1a-d, I will first discuss three data-related suggestions that fall outside these topics.

### 3.1.0 TOR1: Three data-related suggestions

I suggest that:

- echo data down to the top of the acoustic dead zone be used in the assessment,
- age-length keys (ALKs) be used more consistently, and
- consideration be given to random checking of the decision to read otoliths whole

Only the first of these suggestions might have a significant effect on the assessment; the other two are refinements to existing procedures.

*Integration range of the acoustic-trawl data*

Currently, although echoes recorded in the acoustic-trawl survey (ATS) are integrated down to the top of the acoustic dead zone, only the portion more than 3 m above the bottom is used in the assessment. The original rationale for this was that the bottom-trawl survey (BTS) was already covering "near-bottom" pollock (i.e., those below the 3 m nominal headline height of the BTS net) so it was decided that the ATS data in the assessment would be restricted to pollock further from the bottom. There are two reasons to reconsider this decision: recent research has shown that the effective headline height of the BTS net can be substantially greater than 3 m (Kotwicki et al. 2015); and it is wasteful to discard valuable ATS information about near-bottom fish. Since the BTS and ATS operate independently (and are treated as independent in the assessment), there is no reason for concern about double counting of fish in the near-bottom zone.

*More consistent use of ALKs*

I was surprised at the lack of consistency in the structuring of ALKs in different data sources. Although separate ALKs were constructed east and west of 170°W for both the fishery (AR p. 58) and the ATS (Honkalehto et al. 2009; Woillez et al. 2016), we were told that this was not true for the BTS (with the reason being that it appeared to make little difference).

It is understandable that different people should reach different conclusions about the need to separate ALKs at 170°W. A presentation during the review (at the request of the panel) of ALK data from fishery catches showed that east-west differences in the ALK data were clearly apparent only when the data were averaged across years. Thus, although there is strong evidence for an east-west difference in age distributions by size, it is possible that the use of separate ALKs east and west of 170°W may not be justified because the difference is too small to be reliably characterised by ALK data from a single year.

I suggest that the following procedure be used to decide whether or not to separate ALKs east and west of 170°W in future assessments (and that the decision should be applied consistently to all data sources). Two runs of the present assessment model should be carried out which differ only in that run 1 uses age composition data compiled without longitude stratification of the ALKs, and run 2 uses data with this stratification. Occam's Razor should apply: that is, longitudinal stratification of ALKs should not be used in future assessments unless run 2 is clearly preferred to run 1. For run 2 to be preferred, it should show (a) a clearly better fit to the composition data, and (b) a clear difference in estimated stock status.

*Checking the decision to read otoliths whole*

The ageing of pollock seems to be more secure than is the case for many fish species: the ageing method has been validated using radiometric methods (Kastelle and Kimura 2006); ageing precision is reasonably high; and good quality-control measures are in place in the ageing process (including re-aging of a random subsample of otoliths).

It has been well established for many species (including pollock) that, although age readings from whole otoliths can be acceptable for some younger fish, accurate ages for older fish (and some young fish with unclear otoliths) must be based on cross sections (from broken and burnt otoliths) or thin sections. For pollock at the Alaska Fishery Science Center, the otolith reader makes a subjective decision as to which otoliths can be read whole, and which from cross sections. This seems to me a reasonable procedure which must increase the cost-efficiency of pollock ageing. However, I wondered if the quality control system might be enhanced if it formally included a check on the decision as to which otoliths are read whole. This would involve (a) ensuring that the random sample of otoliths selected for re-ageing includes a reasonable number that were read whole; (b) re-reading at least some of these otoliths after breaking and burning; and (c) providing feedback to readers about the soundness of their decisions as to which otoliths to read whole.

### 3.1.1 TOR1a: The AVO index
**Is the [use of the] index of acoustic backscatter from opportunistic (AVO) used appropriately?**

I found no reason to doubt that the AVO index (Honkalehto et al. 2014) was used appropriately in the assessment. The fact that it uses only acoustic backscatter data (but no information about the length distribution or target strength values for pollock) validates the decision to treat the AVO as indexing biomass, rather than numbers. The very high correlation (0.95) between the AVO biomass index and the corresponding (but completely independent) ATS biomass indices (in AR Table 1.19) provides some assurance that AVO indices are useful, particularly in years when there was no ATS (though, given the small sample size [$n$ = 7], I would caution against putting too much weight on the exact value of this correlation).

I discuss in Section 3.1.4 how the weighting of this index in the assessment could be improved.

### 3.1.2 TOR1b: Numbers vs biomass
**Is modeling observed numbers from surveys appropriate?**

I understand this question to relate to the decision that stock assessment scientists make when deciding whether a survey abundance index should be presented to the assessment model as an observation of fish numbers or biomass. Of the four survey indices used in the assessment, three (BTS, ATS-age-1, ATS-age-2+) were presented as numbers, and one (AVO) as biomass. I will restrict my comments here to two of these indices, because I have already noted that the decision to treat AVO as biomass was appropriate (see Section 3.1.1), and I explain below why I think the ATS-age-1 index should be removed from the assessment (see Section 3.2.5).

For the ATS-age-2+ and BTS indices, it would seem sensible to make the decision on whether to use biomass or numbers on the basis of which quantity is more precisely estimated in each survey. Preliminary information suggests that it would have been better to have used biomass, rather than numbers, for these indices. For the ATS, Woillez et al. (2016) found that biomass estimates were more precise, and sometimes substantially so (see the bottom row of their table 4, but note that these calculations included age 1 fish). During the review meeting Stan Kotwicki offered the opinion that biomass estimates would also be more precise for the BTS.

If either or both of these two indices (ATS-age-2+ and BTS) are presented as biomass in future assessments, it will be important to input to the assessment model survey estimates of mean body mass at age by year so that the model can properly convert estimates of expected numbers at age to biomass.

### 3.1.3 TOR1c: Body mass at age
**How should data on mean body mass at age be best used for model projections?**

A new feature of the assessment model in 2015 was the use of a growth model (fitted to observed fishery mean body mass at age data) to predict the mean body mass at age in 2015-2017, and calculate the uncertainty associated with these predictions. These predictions, and their associated uncertainties, were intended to improve the assessment projections by (a) providing better estimates of mean body mass in 2015-2017, and (b) allowing uncertainty about these predictions to propagate into the calculations of $F_{\mathrm{MSY}}$ (AR p. 63).

The idea of this new development was good; the growth model generally fitted the data well and so ought to be useful in predicting body mass for 2015-2017. However, because of an oversight in its implementation, the model did not produce sensible predictions. This is evident in the top right panel of AR Fig. 1.16, where it can be seen that, for all but the youngest cohort, the predicted 2015 mean body mass is less than that in the previous year, implying that fish shrank (e.g., the table says that the mean body mass of the 2008 cohort shrank from 0.74 kg [at age 6 in 2014] to 0.67 kg [at age 7 in 2015]). As a simple way to estimate the effect of this error, I compared the predicted 2015 mean body masses from AR Fig. 1.16 with those I calculated by applying mean age-specific percentage growth rates (calculated from either the observed or estimated mass values in AR Fig. 1.16) to the estimated 2014 mean body masses (Fig. 1) and found that the predicted values for 2015 were too low by on average 22% (i.e., the black line in Fig. 1 is about 22% less than the red and green lines). The predicted body masses for 2016 and 2017 are also likely to be too low by about 22%.

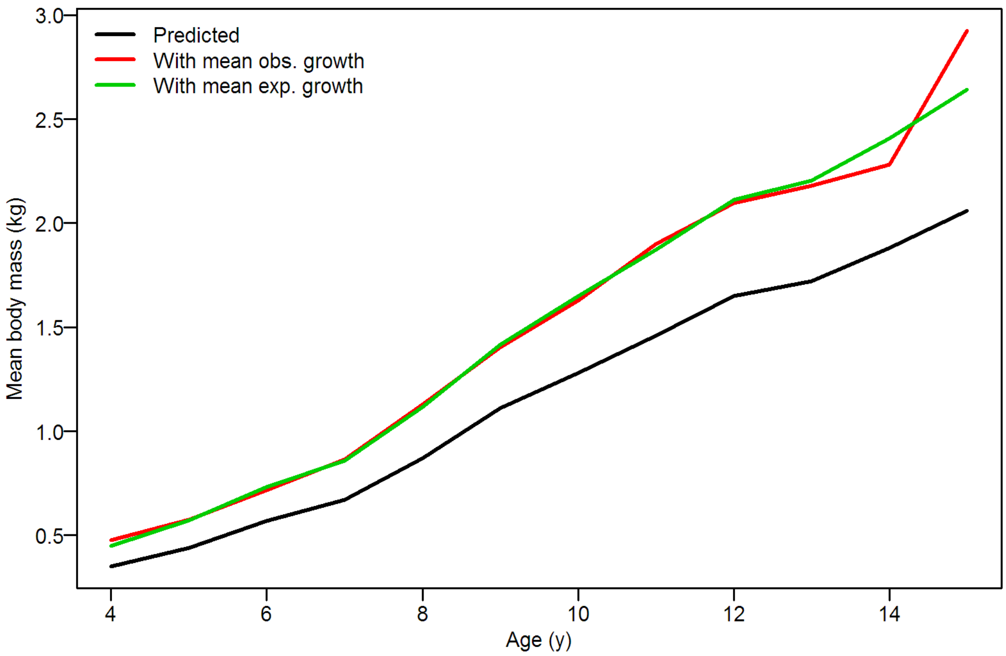


**Figure 1: Comparison of the predicted 2015 mean body mass at age (as shown in AR Fig. 1.16, and used in projections) with values calculated by applying mean observed or expected percentage growth rates by age to the 2014 estimated mean body mass at age.**

This error is potentially serious. I think it means that projected catches and biomasses are likely to have been about 22% too low, but I haven't thought through the effect of this on harvest recommendations.

I think it should not be too difficult to correct the error. During the review, Jim Ianelli suggested this was simply a matter of modelling annual age-specific growth in body mass, rather than mean body mass at age. That seems plausible to me.

### 3.1.4 TOR1d: Data weighting
**How should the various data sets be weighted?**

As I have noted elsewhere (Francis 2011), the weighting of data sets in a stock assessment model is important for two reasons: it can substantially affect the results of a stock assessment; and inappropriate weighting invalidates estimates of uncertainty (e.g., CVs of model outputs such as $F_{\mathrm{MSY}}$). The following brief comments are based on the detailed approach to data weighting that I outline in the paper just cited.

Two types of data required weighting in this assessment: abundance indices (indexing either biomass or numbers of fish), and age compositions. The former were weighted using coefficients of variation (CVs) [I am following the AR in calling these weights CVs; technically, for those indices with lognormal likelihoods, these weights were actually log-space SDs (standard deviations), which are not quite the same as CVs: CV = $(\exp(\text{log-space-SD}^2) - 1)^{0.5}$], and the latter using effective sample sizes ($N$s) (Table 1). Ideally, these weighting parameters should be set to be consistent with the variance of the associated residuals, $V(O_i - E_i)$, where, for the $i$th datum, $O_i$ is the observed value and $E_i$ is the value expected by the model. It is important to understand that this variance has two independent components – $V(O_i - E_i) = V(O_i - T_i) + V(T_i - E_i)$ (where $T_i$ is the true (real-world) value of the observed quantity) – the first term representing observation error and the second what Francis (2011) called process error (which arises because the stock assessment model is only an approximation to the real world).

**Table 1: Summary of the weights applied to the two data types in the assessment, together with two statistics designed to evaluate the consistency of these weights with the model residuals: SDNR (standard deviation of the normalised residuals) for the abundance indices, and the factor, *w*, from applying reweighting algorithm TA1.8 to the composition data. See text for details.**

| Abundance indices | BTS | ATS-age-2+ | CPUE | AVO | ATS-age-1 |
|---|---|---|---|---|---|
| Median CV | 0.15 | 0.20 | 0.20 | 0.27 | 1.0 |
| SDNR | 1.95 | 1.52 | 0.39 | 0.77 | 3.4 |

| Age compositions | Fishery(post-1990) | BTS | ATS | Fishery(1964-1977) | Fishery(1978-1990) |
|---|---|---|---|---|---|
| Median *N* | 325 | 100 | 50 | 50 | 10 |
| *w* | 1.6 | 1.7 | 0.54 | 1.8 | 1.8 |

I think there is room for more rigour in the weighting of the abundance indices. For BTS, the approach used to calculate CVs seems sound, but these CVs represent only observation error; I suggest adding a CV of 0.2 (see Francis et al. 2003) to allow for process error (e.g., temporal variation in catchability) [note that CVs add as squares so, with an observation-error CV of 0.15, say, the total CV would be $0.25 = (0.15^2 + 0.20^2)^{0.5}$]. I won't discuss the weighting of the ATS-age-1 index because I don't believe this index should be used in the assessment (see Section 3.2.5). The current weighting of the other two acoustic indices (ATS-age-2+ and AVO) involved calculating CVs for one error component (variability in acoustic backscatter, quantified using geostatistics) and scaling up these CVs by some multiplier. This has two weaknesses: first, the scaling multiplier is arbitrary; second, we can't expect the total CV for these surveys to be proportional to the backscatter CV (this is well illustrated in table 5 of Woillez et al., 2016, which shows that, over five surveys, the relative contribution of backscatter to total uncertainty in biomass ranged from 26 to 82). For these acoustic indices I suggest a simulation approach that attempts to quantify all sources of uncertainty, including both observation and process error (e.g., Rose et al. 2000; O'Driscoll 2004). The details of such procedures are outside my area of expertise but the approach of Woillez et al. (2016) seems promising. Finally, for the CPUE data I suggest using a data smoother to calculate an appropriate CV, following Clark and Hare (2006, p. 9) (see Francis, 2011 for a rationale supporting this approach).

It is common practice to calculate SDNRs (standard deviations of the normalised residuals) as a model diagnostic for abundance indices. The important thing to understand about these SDNRs is that they should *not* be used to iteratively reweight the abundance data (Francis 2011). For the current assessment the SDNRs suggest that residuals were, on average, smaller than expected for CPUE and AVO, and larger than expected for the other indices (especially ATS-age-1) (Table 1). However, it is hard to interpret these particular SDNRs because, as noted in the previous paragraph, I think insufficient rigour was applied in deriving the CVs on which they depend. When CVs are more rigorously derived, an SDNR much greater than 1 suggests that the associated abundance index is not well fitted, and our focus should be on seeing whether we can adjust our model assumptions to ensure a better fit. To this end, one question to ask is whether other, non-abundance, data might be over-weighted.

The SDNR diagnostic needs to be modified for the BTS index because the multivariate normal likelihood used for this index includes correlations between years. I believe that the appropriate statistic is $\left[(\boldsymbol{r}^T\mathbf{V}^{-1}\boldsymbol{r})/n_{year}\right]^{0.5}$, where $\mathbf{r}$ is the (column) vector of residuals, $^T$ denotes the matrix transposition operation (so $\mathbf{r}^T$ is a row vector), $\mathbf{V}$ is the covariance matrix of the multivariate normal, and $n_{\text{year}}$ is the number of years in the index. Note that this statistic becomes an SDNR in the special case when all correlations are zero (so that $\mathbf{V}$ is a diagonal matrix).

The main weakness in the weighting of the composition data in the current assessment is that the iterative reweighting algorithm (AR top of p. 144) [which appears to be the same as method TA1.8 of Francis (2011)] has been insufficiently used. It should be used for all composition data sets, and should be reapplied every time the assessment is substantially changed, either by the addition of new data or changes in assumptions. My understanding is that it was last applied in the EBS pollock assessment several years ago, and then it was not applied to all compositions. When applied to outputs from the current assessment it suggested that *N*s should be lower for ATS, and higher for all other data (see $w$ values in Table 1, noting that, for each data set, $N_{y,\mathrm{new}} = w\, N_{y,\mathrm{input}}$, where $y$ indexes the years in the data set). A good way to calculate input sample sizes (i.e., $N_{y,\mathrm{input}}$) for composition data is to quantify the observation error by bootstrapping the raw data from which each composition was constructed. I was pleased to see that this was done (though not, unfortunately, in the current assessment) for the recent fishery compositions, and suggest that it also be done for the compositions from the surveys and earlier fisheries. Because of the strong between-age correlations that are common in bootstrap samples of compositions, the calculation of the $N_{y,\mathrm{input}}$ should be based on the variability of mean age, rather than of the individual proportions at age (which I think was used for the current assessment). The calculation is simple: $N_{y,\mathrm{input}} = v_y / \mathrm{V}_j(\bar{A}_{jy})$, where $\bar{A}_{jy}$ is the mean age for the $j$th bootstrap composition in year $y$, $\mathrm{V}_j$ denotes the variance over the bootstrap samples, $v_y$, the variance of age in the observed composition in year $y$, is calculated as $v_y = \sum_a (a^2 p_{ay}) - \left(\sum_a a p_{ay}\right)^2$, and $p_{ay}$ is the proportion at age $a$ in the observed composition for year $y$. (I discuss below how $N_{y,\mathrm{input}}$ values calculated in this way can be used to test whether composition data are over-fitted – see Section 3.2.1).

Finally, a note of caution on the interpretation of the SDNR and $w$ statistics of Table 1. Both statistics are often quite uncertain because their calculation uses variances from samples that are often quite small (in both cases the sample size is $n_{\mathrm{year}}$, the number of years in the data set). For example, the SDNR of 0.77 for AVO (for which $n_{\mathrm{year}} = 10$) is well within the 95% confidence interval for the SD of a sample of size 10 from a population with true SD = 1. It is easy to calculate 95% confidence intervals for both the SDNR and $w$ statistics by resampling residuals (see function SSMethod.TA1.8 in R package r4ss [Taylor et al. 2014] for how this can be done in calculating $w$). For the ATS $w$, the point estimate is 0.54, but its 95% confidence interval is (0.36,1.05). I am *not* suggesting a hypothesis-testing approach to these statistics in which we test a null hypothesis that the true value is 1 and take action only when this hypothesis is rejected. I am simply suggesting that we take the uncertainty of the statistics into account when interpreting, and acting upon, them.

## 3.2 TOR 2: Assessment methods

**Evaluate and provide recommendations on model structure, assumptions, and estimation procedures uses to assess stock status and condition.**

My overall impression is that this is a high quality assessment, generally using state-of-the-art model structures and assumptions.

### 3.2.1 TOR2a: Selectivity

**Are the selectivity approaches used for surveys and fishery appropriate?**

I am not able to answer to this question directly. The approaches are complex, involving three quite different ways of implementing time variation in selectivity (for the fishery, BTS, and ATS) and *many* assumptions (including both the choice of parametric forms [e.g., AR Eq. 4, p. 143] and the fixing of variances for random deviates [e.g., $\sigma_E^2$ and $\sigma_S^2$ on AR p. 142] and weights for penalties [e.g., the "curvature" penalty of AR p. 142]). To evaluate these approaches, I would need to see the evidence supporting all these assumptions. I imagine this evidence has accreted gradually as a result of experimentation over a number of years, so it would not be simple to assemble it and present it in an intelligible manner. I think I understand why time variation in the fishery selectivity was made to be much more flexible than that for the two surveys, but I have no way of judging whether the chosen degree of flexibility was appropriate in each case; and I do not know the criteria by which the appropriate degree of flexibility was decided. Nor do I know whether the BTS selectivity should require so many more parameters than that for the ATS.

In lieu of a direct answer to the question posed in this TOR, I offer:

- a way to provide the assessment model with more information about variation in selectivity,
- a method to evaluate competing approaches to selectivity, and
- a way of testing whether over-flexible selectivities are causing over-fitting of composition data

I think it is worth investigating the effect of providing the assessment model with more (and more direct) information than it currently has about temporal changes in fishery selectivity. The additional information could improve the assessment by increasing the precision of the selectivity estimates, and thus the precision of recruitment estimates. The main way to do this is to disaggregate the fishery composition data (and catches) into sub-fisheries which are defined in such a way as to minimize differences in age compositions within the same sub-fishery, and maximise differences between sub-fisheries. Techniques like tree-based regression are well suited to identify the variables (e.g., vessel or gear type, month, time of day, bottom depth, area) which best predict some composition statistic (e.g., mean age, proportion of young fish, etc.) and the break points in those variables that minimise within-sub-fishery variation. The overall fishery selectivity for each year is the weighted sum of the sub-fishery selectivities (with the weights being the share of the catch caught by each sub-fishery) and much of the year-to-year variation in the overall fishery selectivity is likely to be caused by changes in these weights (which are the additional information I refer to in the first sentence of this paragraph). Given the wealth and quality of the fishery composition data for this assessment, I found it unusual, and surprising, that there was no disaggregation of the fishery data. A second way to provide more information about selectivity is to identify years in which there were significant changes in either management or fishing practice that are likely to induce step changes in selectivity, and to build these step changes into the selectivity model.

Rather than asking whether the current selectivity assumptions are appropriate, I suggest that it is easier (though not always easy) to ask which of two (or more) approaches to selectivity is better. This is particularly so when (as is desirable) the approaches are nested (i.e., one is a simplification of the other). In this case, Occam's Razor helps us because it provides a default answer: when one model is not clearly superior to the other we should choose the simpler model. There are two criteria that are relevant in judging whether a more complex model is superior: does it provide a clear improvement in fit to the non-composition data? (goodness of fit to the composition data is irrelevant to this judgement – adding selectivity parameters [or increasing random-effects variances] virtually always improves the fit to compositions); and would it lead to markedly different management advice? It's worth remembering that our aim in stock assessments is to provide the best management advice; it is not to provide the most realistic population model. Models that are more complex, and thus might be deemed more realistic, sometimes provide less robust advice. Another point to remember is that any comparison between competing models must be fair to both models. At one stage during the review meeting, we were told about the effect on the assessment results of turning off time-varying selectivity. This was not a fair comparison with the base model because it is typically the case that base models are decided on only after a careful examination of all model diagnostics to ensure that no assumptions need be tweaked to improve these diagnostics. For a fair comparison, a similar examination needs to be made for any alternative model (in the case of the fixed-selectivity model this should have included reweighting the composition data). I mention this fixed-selectivity model simply to make my point about fair comparisons. When evaluating the approach to selectivity, I think it's probably not useful to comparing models that differ so greatly in complexity (the base model has hundreds more parameters).

My aim in suggesting the evaluations described in the previous paragraph is to provide a response to the valid criticism by Nielsen and Berg (2014) that the structuring of time-varying selectivity in conventional penalised-likelihood models is typically *ad hoc*. This seems particularly true of fixed variances of random deviates, but also concerns other decisions described above. By demonstrating how the chosen selectivity structure is superior to alternative structures, we make this part of our assessment model less *ad hoc*.

A clear danger associated with time-varying selectivities is the possibility of over-fitting the composition data. This is occurring if parts of the year-to-year changes in a selectivity are in response to sampling noise in these data, rather than changes in selectivity. I suggest that there is clear evidence of over-fitting of a composition data set if $\text{median}_y(N_y) > \text{median}_y(N_{y,\text{input}})$, where, for year $y$, $N_{y,\text{input}}$ is the observation-error effective sample size (calculated by bootstrapping, as described in Section 3.1.4), and the $N_y$ are the effective sample sizes used in the assessment for this data set. The rationale behind this suggestion is straightforward. As noted above, the variance of the total error in the composition data is the sum of two independent components: $V_{\text{total}} = V_{\text{observation}} + V_{\text{process}}$. When we introduce time-varying selectivity into an assessment, we reduce the process error associated with a time-invariant selectivity. Since variances must be positive, we will clearly have gone too far (i.e., be over-fitting the data) if $V_{\text{total}} < V_{\text{observation}}$. Since these variances are, for the multinomial likelihood, inversely proportional to the effective sample sizes, this is equivalent to $\text{median}_y(N_y) > \text{median}_y(N_{y,\text{input}})$. It should be noted that this test is conservative. That is, the fact that over-fitting is not detected (i.e., $\text{median}_y(N_y) < \text{median}_y(N_{y,\text{input}})$) does not mean that it is not occurring. The point is that time-varying selectivity removes only one type of process error (that from wrongly-specified selectivities); it cannot remove other types (e.g., from wrongly-specified natural mortality).

### 3.2.2 TOR2b: Trans-boundary issues
**How should trans-boundary aspects of the resource be handled?**

It is very common for assessment scientists to decide to ignore evidence of fish movement across assumed stock boundaries because (a) the extent of movement seems small compared to the stock size, and (b) they have no practical way to include the movement in the assessment. Such a decision was made in the present assessment with regard to fish movements across the US-Russian boundary at the north-western edge of the assumed EBS stock area. The information I saw gave me no cause for concern about this decision.

### 3.2.3 TOR2c: Survey catchability
**What constraints, if any, should be placed on survey catchability?**

There are two types of constraints that could be applied to survey catchabilities in this assessment, though I think it possible that neither would have a substantial effect. The first is in the form of prior distributions informed by expert judgement, which have been used for both trawl and acoustics surveys in New Zealand hoki assessments. These were constructed by assigning likely bounds to each of the multiplicative factors which might affect survey catchability (e.g., horizontal herding for a trawl survey) and simulating a prior distribution by sampling each factor from within its bounds. The resultant lognormal priors for hoki surveys were quite broad (CVs ranged from 0.61 to 0.77; see table 28 and figure 52 in Francis, 2009).

A second possible constraint is a prior on the ratio of the two BTS catchabilities (one for 1983, 1984, and 1986, when only strata 1-6 were covered, and the other for the remaining years) which could be constructed from the biomass estimates and CVs from the years in which all strata were occupied. I note that data in AR Table 1.11 suggest that this prior would have a mean of about 0.95, which is quite different from the estimate of 0.43 in the current assessment (I calculated this from the parameter estimates of AR Table 1.22, assuming that the two catchabilities were q_std_area and q_bts, respectively). Catchability-ratio priors like this are implemented in CASAL (for details see section 6.7.5 of Bull et al., 2012).

### 3.2.4 TOR2d: Projections
**How should model projection alternatives be evaluated/presented?**

The set of seven projections carried out for this assessment were designed to meet the requirements of three separate documents: Amendment 56 of the BSAI Groundfish Fishery Management Plan; the National Environmental Policy Act; and the Magnuson-Stevens Fishery Conservation and Management Act. It is beyond both the scope of this review and my expertise to judge whether these projections, and the manner in which they were presented, satisfy these requirements. However, I will make a comment about uncertainty.

I note that, although considerable effort was made to propagate assessment uncertainties into these projections, the results would have under-estimated uncertainty because they were based on a single model, and thus did not capture uncertainty about the assumptions of that model (e.g., about natural mortality). However, this may not matter because, as far as I could tell, where the projection results were used (e.g., in status determination – see AR pages 70-71), projection uncertainty was ignored.

### 3.2.5 TOR2e: Other matters
**Anything else on which the reviewers care to comment.**

I will comment on two other matters: the estimation of steepness and the ATS-age-1 index.

*Estimation of steepness*

I think that the estimation of steepness in this assessment was ill-conceived. We were told that initial unconstrained estimates of steepness were implausibly high and that a more plausible estimate was obtained by (a) arbitrarily increasing the recruitment variability parameter, $\sigma_R$, to 0.9, and (b) using a "conservative prior" (AR p. 145) for steepness. These two steps seem fundamentally wrong to me. First, $\sigma_R = 0.9$ is inconsistent with model estimates of recruitment, which I understand imply a value of about 0.67. Second, and more importantly, what was called a prior was actually an arbitrary penalty. I infer this from both the above-quoted use of "conservative" and the fact that the estimated steepness of 0.68 is three prior standard deviations above the prior mean (assuming that the prior mean and CV given on AR p. 145 are correct). If this prior were really an expression of belief about the true value of steepness we should have to conclude that the data are, in the context of the model assumptions (particularly the stationarity of the stock-recruit relationship), inconsistent with this prior belief and so do not allow an estimate of steepness. It is grossly misleading to characterise our uncertainty about EBS pollock steepness by presenting a point estimate (0.68) and standard error (0.07) (AR Table 1.22) that are driven primarily by these two arbitrary assumptions, rather than by the data. A more honest approach would be to fix steepness at an acceptable level for the base runs and evaluate uncertainty using sensitivity runs using bracketing values of steepness. There is no shame in acknowledging an inability to estimate steepness; it is notoriously difficult to estimate, and few assessments will have sufficient data to do so (Lee et al. 2012).

*The ATS-age-1 index*

I think this index should be excluded from the assessment because it is so uncertain that it can make no material contribution. I note that an SDNR of 3.4 (Table 1) implies that the already *very* high assumed CV of 1.0 for this index may be too low by a factor of more than 3.

## 3.3 TOR 3: Harvest recommendations

**Evaluate and provide recommendations on harvest recommendations provided by the NPFMC Tier system in the context of the 2,000,000 t BSAI cap and realized management recommendations.**

The time available for this review did not allow a thorough evaluation of the NPFMC tier system for providing harvest recommendations. However, from what I understand, it appears to be consistent with best international practice. It is both common and sensible to devise multi-tier management systems to allow for the fact that how we manage a fishery should depend on the level of information we have about the stock, and this level will vary widely between stocks. The only comment I would make concerns one criterion.

The criterion for inclusion in Tier 1 – "reliable estimates of $F_{\mathrm{MSY}}$ and its pdf [probability density function]" (AR p. 69) – seems unrealistically stringent. Given the quantity and quality of available data, and the quality and rigour of the assessment, it is clear that EBS pollock must belong to the top tier (Tier 1) of any multi-tier system. However, I don't believe that the present assessment meets the above criterion. $F_{\mathrm{MSY}}$ depends strongly on (*inter alia*) two demographic parameters: natural mortality and steepness. Thus, any uncertainty about these parameters will affect the pdf of $F_{\mathrm{MSY}}$. Given that the former parameter was fixed in this assessment and the estimated uncertainty of the latter was "grossly misleading" (see Section 3.2.5), I don't see how we could conclude that the current estimate of the uncertainty of $F_{\mathrm{MSY}}$ (as described by the inferred pdf) is "reliable". (I am also a little concerned about the assumption that we can infer the pdf of $F_{\mathrm{MSY}}$ from its point estimate and Hessian-derived standard error with an assumed lognormal distribution – but that is a secondary concern).

## 3.4 TOR 4: Use of ecosystem data

**Evaluate the extent that ecosystem data are presently included in the assessment and recommend how and where improvements can be made.**

The inclusion of ecosystem considerations in fisheries assessments is a complex topic in which I have little experience or expertise. However, I was impressed both by the extent to which ecosystem data and considerations are currently included in Alaskan fisheries management (e.g., via the Ecosystems Considerations report), and by what we were shown of active research in the development of relevant tools (e.g., the CEATTLE multi-species stock assessment model and the Alaska Climate Change Integrated Modeling project).

## 3.5 Documentation

I was disappointed to find that my ability to review this high-quality assessment was compromised by the low-quality documentation of model assumptions. I understand that assessment reports are not highly polished documents. They usually contain many (usually minor) typographical and other errors that are inevitable consequences of the time constraints under which they are prepared. However, the report for this assessment was, in my experience, unusually poor. The main problem was that the documentation has not kept up with the many changes in assumptions that have occurred as the assessment has changed over time. I was misled by the fact that some model features that are no longer used are still described as if they were (e.g., the prior on the sum of the BTS and ATS catchabilities [AR p. 64] and the environmental term $\kappa_t$ in Eq. 8) and some changes in fixed parameters were not recorded and explained (e.g., age composition sample sizes for ATS in AR Table 1.21, and the CV assumed for the ATS-age-1 index). Also, I wasted some time before the review meeting puzzling over the (erroneous) description of the EBS stock area as being "west of 170° W longitude" (AR Appendix 1.1 p. 1), and found during the meeting that, contrary to what is implied by AR Eq. 20, AVO indexed biomass, rather than numbers (this was important for TOR 1b).

Many of the inadequacies of the documentation would have been of much less consequence had the report included (as is increasingly common in assessment reports), appendices containing the model code (i.e., the tpl file) and the input files for the base model (I did ask for the tpl file during the review meeting but was told that it was not in a fit state to be read by others). Even when assessment authors have been careful about keeping their model descriptions up to date, I have learned to be sceptical about equations in reports because it is very easy to write an equation that does not accurately represent what is in the model code. Thus having access to the program code is important. I was grateful for the list of model parameters provided in AR Table 1.22, because this alerted me to some deficiencies in the documentation, so that I could ask for clarification. Although Jim Ianelli was very ready to respond to such queries during the review meeting, I was left with the uneasy feeling that there may be aspects of the model that I have misunderstood or am still unaware of.

# 4. Conclusions and recommendations

The EBS pollock assessment is impressive. It uses an unusually extensive array of generally high-quality data in a well-structured model. The only obvious weaknesses related to an historical decision about the vertical extent of echoes included in the ATS survey data and an oversight in the structure of a new body-mass growth model.

Of the following recommendations, only the first two are of particular importance. Recommendations 3-9 should be considered as refinements to the assessment (or data preparation) which may or may not have much effect; recommendation 10 concerns a refinement to the Tier system; and recommendation 11 is intended to make the task of future reviewers more easy.

I recommend that:

1. The index from the ATS survey include echoes down to the top of the acoustic dead zone (Section 3.1.0),
2. The body-mass growth model be modified to produce more reliable predictions of future values (Section 3.1.3),
3. The use of age-length keys in constructing age composition data be made more consistent (Section 3.1.0),
4. Consideration be given to extending the current otolith-reading quality-control procedures to include an explicit check of the readers' decision to read some otoliths whole (Section 3.1.0),
5. The BTS and ATS-age-2+ surveys be treated as indexing biomass if it is confirmed that these surveys estimate biomass more precisely than numbers of fish (Section 3.1.2),
6. The weighting of data inputs in the assessment model be done more rigorously (Section 3.1.4),
7. The current approach to selectivity be evaluated by (a) formally comparing it to alternative approaches to ensure that it is adequately flexible but not excessively complex; (b) investigating whether it might be improved by disaggregating the fishery data (to provide more direct information about year-to-year changes in selectivity); and (c) testing to ensure that the composition data are not over-fitted (Section 3.2.1),
8. The stock-recruit steepness parameter be fixed in the assessment model (Section 3.2.5),
9. The ATS-age-1 index be dropped from the assessment (Section 3.2.5),
10. Consideration be given to making the criterion for a fish stock's inclusion in Tier 1 less unrealistically stringent (Section 3.3), and
11. The assessment documentation be improved by (a) ensuring that it is consistent with the current assessment, and (b) including as appendices the model code (i.e., the tpl) and input files (Section 3.5).

## 5. References


Bull, B., Francis, R.I.C.C., Dunn, A., McKenzie, A., Gilbert, D.J., Smith, M.H., Bian, R., Fu, D., 2012. CASAL (C++ algorithmic stock assessment laboratory): CASAL User Manual v2.30-2012/03/21. *NIWA Technical Report 135*. 280 p. Available from http://www.niwa.co.nz/fisheries/tools-resources/casal.

Clark, W.G., and Hare, S.R. 2006. Assessment and management of Pacific halibut: data, methods, and policy. *Scientific Report 83*. International Pacific Halibut Commission, Seattle, Wash.

Francis, R.I.C.C. (2009). Assessment of hoki (*Macruronus novaezelandiae*) in 2008. *New Zealand Fisheries Assessment Report 2009/7*. 80 p. Available from http://fs.fish.govt.nz/Page.aspx?pk=113&dk=21882.

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences 68*: 1124–1138.

Francis, R.I.C.C.; Hurst, R.J.; Renwick, J.A. (2003). Quantifying annual variation in catchability for commercial and research fishing. *Fishery Bulletin 101*: 293–304.

Honkalehto, T.; Jones, D.; McCarthy, A.; McKelvey, D.; Guttormsen, M.; Williams, K.; Williamson., N. 2009. Results of the echo integration trawl survey of walleye pollock (*Theragra chalcogramma*) on the U.S. and Russian Bering Sea shelf in June and July 2008. U.S. Department of Commerce, NOAA *Technical Memorandum NMFS-AFSC-194*, 56 p.

Honkalehto, T.; Ressler, P.H.; Stienessen, S.C.; Berkowitz, Z.; Towler, R.H.; McCarthy, A.L.; Lauth, R.R. 2014. Acoustic Vessel-of-Opportunity (AVO) index for midwater Bering Sea walleye pollock, 2012-2013. *AFSC Processed Report 2014-04*, 19 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

Kastelle, C.R.; Kimura, D.K. 2006. Age validation of walleye pollock (*Theragra chalcogramma*) from the Gulf of Alaska using the disequilibrium of Pb-210 and Ra-226. *ICES Journal of Marine Science 63*: 1520–1529.

Kotwicki, S.; Horne, J.K.; Punt, A.E.; Ianelli, J.N. 2015. Factors affecting the availability of walleye pollock to acoustic and bottom trawl survey gear. *ICES Journal of Marine Science 72(5)*: 1425–1439.

Kotwicki, S.; Ianelli, J.N.; Punt, A.E. 2015. Correcting density-dependent effects in abundance estimates from bottom-trawl surveys. *ICES Journal of Marine Science 71(5)*: 1107–1116.

Nielsen, A.; Berg, C.W.; 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research 158*; 96–101.

Lee, H-H.; Maunder, M.N.; Piner, K.R.; Methot, R.D. 2012. Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? *Fisheries Research 125-126*: 254–261.

O'Driscoll, R.L. 2004. Estimating uncertainty associated with acoustic surveys of spawning hoki (*Macruronus novaezelandiae*) in Cook Strait, New Zealand. *ICES Journal of Marine Science 61(1)*: 84–97.

Rose, G.; Gauthier, S.; Lawson, G. 2000. Acoustic surveys in the full monte: simulating uncertainty. *Aquatic Living Resources 13(5)*: 367–372.

Taylor, I.; Stewart, I.; Hicks, A.; Garrison, T.; Punt, A.; Wallace, J.; Wetzel, C.; Thorson, J.; Takeuchi, Y.; Monnahan, C. 2014. Package r4ss. https://github.com/r4ss.

Woillez, M.; Walline, P.D.; Ianelli, J.; Dorn, M.W.; Wilson, C.W.; Punt, A.E. 2016. Evaluating total uncertainty for biomass- and abundance-at-age estimates from eastern Bering Sea walleye pollock acoustic-trawl surveys. *ICES Journal of Marine Science* (advanced access online publication).

# Appendix 1 Materials Provided

Before the review workshop, extensive materials, too numerous to list here, were provided via a dedicated website (http://www.afsc.noaa.gov/REFM/Stocks/plan_team/2016EBSpollockCIE). These included the assessment report (with a separate appendix on stock structure), reports from previous CIE reviews (of the assessment in 2010, and of recruitment processes in 2015), and background documents on management matters, the observer program, surveys, and ecosystem issues. Most presentations made during the review were also made available via the same website.

# Appendix 2: Statement of Work

This appendix contains the Statement of Work, including two annexes, that formed part of the consulting agreement between Northern Taiga Ventures Inc. and the author.

**External Independent Peer Review by the Center for Independent Experts**

**Assessment of the pollock stock in the Eastern Bering Sea**

**Scope of Work and CIE Process:** The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

**Project Description:** The annual assessments of the pollock stock in the EBS have used similar model configurations for a number of years now. Review is needed to identify areas where the assessment can be improved and aspects that would lead to best-practices for near term catch recommendations. The SSC has requested evaluation of environmental covariates for relative cohort strength, and temperature effects on survey catchability and/or selectivity. Other evaluations on the effect of alternative catch scenarios (i.e., if the catch was equal to the ABC) would be useful to help provide context to the current management practices (in which catches are in most years constrained by a 2 million t limit for all groundfish in the BSAI region). The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

**Requirements for CIE Reviewers:** Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge and recent experience in the application of stock assessment methods in general, and preferably Stock Synthesis in particular. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

**Location of Peer Review:** Each CIE reviewer shall conduct an independent peer review during the panel review meeting ***scheduled in Seattle, WA during May 16-19, 2016 (or one of the subsequent weeks).***

**Statement of Tasks:** Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number,

country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/
http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or online) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Assessment of the walleye pollock stock in the eastern Bering Sea (~100 p.), including a stock structure evaluation provided as an appendix)
CIE review of the recruitment processes group conducted June 2015
Comments on the final 2015 EBS pollock assessments by the Plan Team and SSC

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs cannot be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

The review meeting will include three main parts:

1. A series of presentations with follow-up questions and discussions by CIE reviewers, to be chaired by an AFSC scientist.
2. Any real-time model runs and evaluations conducted in an informal workshop setting, as proposed by the CIE reviewers.
3. Initial report writing by the CIE reviewers, with opportunity for additional questions of the assessment author.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Other Tasks – Contribution to Summary Report: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

**Specific Tasks for CIE Reviewers:** The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting ***scheduled at the Alaska Fisheries Science Center in Seattle, WA during May 16-19, 2016.***

3) Participate at the peer review meeting ***tentatively scheduled at the Alaska Fisheries Science Center in Seattle, WA during May 16-19, 2016*** as specified herein, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than ***June 3, 2016***, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Dr. Manoj Shivlani, CIE Lead Coordinator, via email to mshivlani@ntvifederal.net, and CIE Regional Coordinator, via email to Dr. David Die ddie@rsmas.miami.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

**Schedule of Milestones and Deliverables:** CIE shall complete the tasks and deliverables described in this SoW in accordance with the following ***tentative*** schedule.

| | |
|---|---|
| *April 4, 2016* | CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact |
| *April 25, 2016* | NMFS Project Contact sends the CIE Reviewers the pre-review documents |
| ***May 16-19, 2016*** | Each reviewer participates and conducts an independent peer review during the panel review meeting |
| *June 6, 2016* | CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator |
| *June 20, 2016* | CIE submits CIE independent peer review reports to the COTR |
| *June 27, 2016* | The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director |

**Modifications to the Statement of Work:** This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on changes. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

**Acceptance of Deliverables:** Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

**Applicable Performance Standards:** The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) The CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) The CIE report shall address each ToR as specified in **Annex 2**,
(3) The CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

**Distribution of Approved Deliverables:** Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

**Support Personnel:**


Allen Shimada
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
Allen Shimada@noaa.gov Phone: 301-427-8174

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
mshivlani@ntvifederal.com Phone: 305-968-7136


**Key Personnel:**

NMFS Project Contact:


James Ianelli, Alaska Fisheries Science Center
NMFS/NOAA Building 4
7600 Sand Point Way NE
Seattle WA 98115
Jim.ianelli@noaa.gov

## Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.
2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.
    a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.
    b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.
    c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.
    d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
    e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:
    Appendix 1: Bibliography of materials provided for review
    Appendix 2: A copy of the CIE Statement of Work
    Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

## Annex 2: Terms of Reference for the Peer Review

**Assessment of Walleye Pollock in the Eastern Bering Sea**

1. Evaluation, findings, and recommendations on quality of input data and methods used to process them for inclusion in the assessment. In particular:
    a. Is the use of the index of acoustic backscatter from opportunistic (AVO) used appropriately?
    b. Is modeling observed numbers from surveys appropriate?
    c. How should data on mean body mass at age be best used for model projections?
    d. How should the various data sets be weighted?
2. Evaluate and provide recommendations on model structure, assumptions, and estimation procedures uses to assess stock status and condition. In particular:
    a. Are the selectivity approaches used for surveys and fishery appropriate?
    b. How should trans-boundary aspects of the resource be handled?
    c. What constraints, if any, should be placed on survey catchability?
    d. How should model projection alternatives be evaluated/presented?
    e. Anything else on which the reviewers care to comment.
3. Evaluate and provide recommendations on harvest recommendations provided by the NPFMC Tier system in the context of the 2,000,000 t BSAI cap and realized management recommendations
4. Evaluate the extent that ecosystem data are presently included in the assessment and recommend how and where improvements can be made.

## Annex 3: Tentative Agenda

# CIE Review of the Eastern Bering Sea Walleye Pollock stock assessment

Alaska Fisheries Science Center
7600 Sand Point Way NE, Seattle, WA 98115
May 16-19, 2016
Building 4; Room 2143 (or TBD)

**Review panel Chair/facilitator:** Anne Hollowed (Anne.Hollowed@noaa.gov)
**Lead assessment author**: Jim Ianelli (Jim.Ianelli@noaa.gov)
**Security and check-in**: Jim Ianelli

*Sessions will run from 9 a.m. to 5 p.m. each day, with time for lunch and morning and afternoon breaks. Discussion will be open to everyone, with priority given to the panel and senior assessment author.*

### Monday, May 16

| | |
|---|---|
| Preliminaries: | |
| 0900 Introductions and adoption of agenda | Chair |
| Data sources (current and potential): | |
| 0910 Overview of data types used in the assessments | Jim I. |
| 0920 Catch accounting system and in-season management | AKRO SF Division |
| 0950 Observer program | Observer program |
| 1020 Break | |
| 1030 EBS trawl survey | Stan Kotwicki |
| 1115 Acoustic trawl survey | Chris Wilson |
| 1200 Lunch | |
| 1300 Age determination | Tom Helser |
| 1330 Age composition and mean-weight-at-age data | Jim I. |
| Assessment models: | |
| 1400 Assessment details | Jim I. |
| 1500 Break | |
| 1510 Management background and issues (ToR 3) | Diana Stram (NPFMC) |
| 1610 Ecosystem application in assessment (ToR 4) | TBD |
| 1640 Discussion | Panel |

### Tuesday, May 17

| | |
|---|---|
| 0900 Assessment model review | Jim |
| 1000 Topics as needed, discussion and model clarifications | |
| 1300 Presentation of model updates, further requests and discussions | |
| 1700 Adjourn | |

### Wednesday, May 18

Review of models assigned the previous day
Discussion, real-time model runs
Assignments for models to be presented the following day

### Thursday, May 19

Review of models
Discussion, real-time model runs
Report writing (time permitting)

## Appendix 3: Participants in the Review Meeting

| Name | Organization | e-mail |
|---|---|---|
| Stan Kotwicki | NMFS – AFSC | Stan.Kotwicki@noaa.gov |
| Martin Dorn | NMFS – AFSC | Martin.Dorn@noaa.gov |
| Diana Stram | NPFMC Staff | Diana.Stram@noaa.gov |
| Kirstin Holsman | NMFS – AFSC | Kirstin.Holsman@noaa.gov |
| Darin Jones | NMFS – AFSC | Darin.Jones@noaa.gov |
| James Thorson | NMFS – AFSC | James.Thorson@noaa.gov |
| Ed Richardson | At Sea Processors Assoc. | erichardson@atsea.org |
| Chris Wilson | NMFS – AFSC | Chris.Wilson@noaa.gov |
| Steve Barbeaux | NMFS – AFSC | Steve.Barbeaux@noaa.gov |
| Craig Faunce | NMFS – AFSC | Craig.Faunce@noaa.gov |
| Steve Martell | Sea State | Steve@seastateinc.com |
| Sandi Neidetcher | NMFS – AFSC | Sandi.Neidetcher@noaa.gov |
| Jim Ianelli | NMFS – AFSC | Jim.Ianelli@noaa.gov |
| Alex De Robertis | NMFS – AFSC | Alex.DiRobertis@Noaa.gov |
| Nate Lauffenburger | NMFS - AFSC | Nate.Lauffenburger@noaa.gov |
| Kresimir Williams | NMFS – AFSC | Kresimir.Williams@noaa.gov |
| Robert Lauth | NMFS – AFSC | Bob.Lauth@noaa.gov |
| Anne Hollowed | NMFS – AFSC | Anne.Hollowed@noaa.gov |
| Chris Francis | CIE | chris.francis@clear.net.nz |
| Henrick Sparholt | CIE | henrik.sparholt@gmail.com |
| Patrick Cordue | CIE | plc@isl-solutions.co.nz |